International Journal of Modern Physics: Conference Series
Vol. 1, No. 1 (2010) 1–5

DOI: 10.1142/insert DOI here

World Scientific
www.worldscientific.com

ROUGH CLUSTERING FOR CANCER DATASETS

TUTUT HERAWAN

Database and Knowledge Management Research Group
Faculty of Computer System and Software Engineering
Universiti Malaysia Pahang
Lebuh Raya Tun Razak, Gambang 26300, Kuantan, Pahang, Malaysia
tutut@ump.edu.my

Received (Day Month Year)
Revised (Day Month Year)

Cancer is becoming a leading cause of death among people in the whole world. It is confirmed that the early detection and accurate diagnosis of this disease can ensure a long survival of the patients. Expert systems and machine learning techniques are gaining popularity in this field because of the effective classification and high diagnostic capability. This paper presents the application of rough set theory for clustering two cancer datasets. These datasets are taken from UCI ML repository. The method is based on MDA technique proposed from Ref. 11. To select a clustering attribute, the maximal degree of the rough attributes dependencies in categorical-valued information systems is used. Further, we use a divide-and-conquer method to partition/cluster the objects. The results show that MDA technique can be used to cluster to the data. Further, we present clusters visualization using two dimensional plot. The plot results provide user friendly navigation to understand the cluster obtained.

Keywords: Clustering; Rough set; MDA technique; Cancer datasets.

1. Introduction

In the last years there has been a considerable growth of the amount of biological data available in several domains. The use of clustering algorithms to discover new and useful information in biological data is getting increasing attention lately. Clustering algorithms are considered a powerful tool for the identification of groups and sub-groups in biological data. Clustering algorithms aim to group data consistently, in such a way that the most similar objects belong to the same group or cluster and dissimilar objects are assigned to different clusters. The use of these algorithms allows to detect similar objects in a dataset that could not be easily or efficiently grouped by humans[1]. Clustering a set of objects into homogeneous classes is a fundamental operation in data mining. The operation is required in a number of data analysis tasks, such as unsupervised classification and data summation, as well as segmentation of large homogeneous data sets into smaller homogeneous subsets that can be easily managed, separately modeled and analyzed. Recently, many attentions have been put on categorical data clustering[2-8], where data objects are made up of non-numerical attributes. One of the popular approaches is based on rough set theory[9-11]. The main idea of the rough clustering is the clustering data set is mapped as the decision table. This can be done by introducing a decision attribute and consequently, a divide-and-conquer method can be used to partition/cluster the objects. In previous papers[12-14], we propose a technique for selecting a clustering attribute in categorical data clustering. The proposed technique, is based on the maximum degree of dependency of attribute[15,16]. We have succeeded in showing that the proposed technique is able to achieve lower computational complexity with higher purity as compared to baseline techniques. Cancer is becoming a leading cause of death among people in the whole world. Expert systems and machine learning techniques are gaining popularity in this field because of the effective classification and high diagnostic capability[17]. This paper presents the application of rough set theory for clustering two cancer datasets. The dataset are taken from UCI ML repository[18]. The method is based on MDA technique proposed by[12-14]. The rough attributes dependencies in categorical-valued information systems is used to select clustering attribute based on the maximal degree. Further, we use a divide-and-conquer method to partition/cluster the objects.

The rest of this paper is organized as follows. Section 2 describes fundamental concept of rough set theory. Section 3 describes the Maximum Attributes Dependency (MDA) technique. Experimental results of MDA on Lung Cancer and Wisconsin Breast Cancer (original) datasets are described in section 4. Finally, the conclusion of this work is described in section 5.

2. Rough Set Theory

2.1. *Information System*

The observation that one cannot distinguish objects on the basis of given information about them is the starting point of rough set theory. In other words, imperfect information causes indiscernibility of objects. The indiscernibility relation induces an approximation space MDA of equivalence classes of indiscernible objects. A rough approximating a subset of the set of objects is a pair of dual approximation operator, called a lower approximation and an upper approximation in term of these equivalence classes. Rough sets are defined trough their dual set approximations in Pawlak approximation space[9]. Here, we use the concept of rough set theory in term of reasoning about data containing in an information system[10]. The notion of information system provides a convenient tool for the representation of objects in terms of their attribute values.

The syntax of information systems is very similar to relations in relational data bases. Entities in relational databases are also represented by tuples of attribute values. An *information system*[11] is a 4-tuple (quadruple) $S=(U,A,V,f)$, where $U=\{u_1,u_2,u_3,\cdots,u_{|U|}\}$ is a non-empty finite set of objects, $A=\{a_1,a_2,a_3,\cdots,a_{|A|}\}$ is a non-empty finite set of attributes, $V=\bigcup_{a\in A}V_a$, V_a is the domain (value set) of attribute *a*, $f:U\times A\rightarrow V$ is an information function such that $f(u,a)\in V_a$, for every $(u,a)\in U\times A$, called information (knowledge) function. An information system is also called a knowledge representation systems or an attribute-valued system and can be intuitively expressed in terms of an information table (see Table 1).

Table 1. An information system

U/A	a_1	a_2	$\cdots$	a_k	$\cdots$	$a_{\|A\|}$
u_1	$f(u_1,a_1)$	$f(u_1,a_2)$	$\cdots$	$f(u_1,a_k)$	$\cdots$	$f(u_1,a_{\|A\|})$
u_2	$f(u_2,a_1)$	$f(u_2,a_2)$	$\cdots$	$f(u_2,a_k)$	$\cdots$	$f(u_2,a_{\|A\|})$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
$u_{\|U\|}$	$f(u_{\|U\|},a_1)$	$f(u_{\|U\|},a_2)$	$\cdots$	$f(u_{\|U\|},a_k)$	$\cdots$	$f(u_{\|U\|},a_{\|A\|})$

The time complexity for computing an information system $S=(U,A,V,f)$ is $|U|\times|A|$ since there are $|U|\times|A|$ values of $f(u_i,a_j)$ to be computed, where $i=1,2,3,\cdots,|U|$ and $j=1,2,3,\cdots,|A|$. Note that t induces a set of maps $t=f(u,a):U\times A\rightarrow V$. Each map is a tuple $t_i=\left(f(u_i,a_1),\cdots,f(u_i,a_{|A|})\right)$, where $i=1,2,3,\cdots,|U|$.

Note that the tuple *t* is not necessarily associated with entity uniquely (see Table 1). In an information table, two distinct entities could have the same tuple representation (duplicated/redundant tuple), which is *not permissible* in relational databases. Thus, the concept of information systems is a generalization of the concept of relational databases.
The starting point of rough set theory is the indiscernibility relation, which is generated by information about objects of interest. The indiscernibility relation is intended to express the fact that due to the lack of knowledge it is difficult to discern some objects employing the available information. That means, in general, it is unable to deal with single objects but clusters of indiscernible objects must be considered. Now the notion of indiscernibility relation between two objects can be defined precisely.

2.2. *Set Approximations*

Definition 1. *Two elements* $x,y\in U$ *are said to be B-indiscernible (indiscernible by the set of attributes* $B\subseteq A$ *in S) if and only if* $f(x,a)=f(y,a)$, *for every* $a\in B$.

Obviously, every subset of *A* induces unique indiscernibility relation. Notice that, an indiscernibility relation induced by the set of attribute *B*, denoted by $IND(B)$, is an equivalence relation. The partition of *U* induced by $IND(B)$ is denoted by U/B and the equivalence class in the partition U/B containing $x\in U$, is denoted by $[x]_B$. The notions of lower and upper approximations of a set are defined as follows.

Definition 2. *The B-lower approximation of X, denoted by* $\underline{B}(X)$ *and B-upper approximations of X, denoted by* $\overline{B}(X)$*, are defined by*

$$\underline{B}(X)=\{x\in U \mid [x]_B \subseteq X\} \textit{ and } \overline{B}(X)=\{x\in U \mid [x]_B \cap X \neq \phi\}, \textit{ respectively.}$$

The accuracy of approximation (accuracy of roughness) of any subset $X \subseteq U$ with respect to $B \subseteq A$, denoted $\alpha_B(X)$ is measured by

$$\alpha_B(X)=\left|\underline{B}(X)\right|/\left|\overline{B}(X)\right|,$$

where $|X|$ denotes the cardinality of *X*. For empty set ϕ, we define $\alpha_B(\phi)=1$. Obviously, $0\leq \alpha_B(X)\leq 1$. If *X* is a union of some equivalence classes, then $\alpha_B(X)=1$. Thus, the set *X* is *crisp* with respect to *B*, and otherwise, if $\alpha_B(X)<1$, *X* is *rough* with respect to *B*.

Another important issue in database analysis is discovering dependencies between attributes. Intuitively, a set of attributes *D* depends totally on a set of attributes *C*, denoted $C \Rightarrow D$, if all values of attributes from *D* are uniquely determined by values of attributes from *C*. In other words, *D* depends totally on *C*, if there a functional dependency between values of *D* and *C*. The formal definition of attributes dependency is given as follows.

2.3. *Dependency of Attributes*

The notion of the dependency of attributes in information systems is given in the following definition.

Definition 3. *Let* $S=(U,A,V,f)$ *be an information system and let D and C be any subsets of A. Attribute D is called depends totally on attribute C, denoted* $C \Rightarrow D$*, if all values of attributes D are uniquely determined by values of attributes C.*

In other words, attribute *D* depends totally on attribute *C*, if there exist a functional dependency between values *D* and *C*. Since an information system is a generalization of a relational database. We would need also a generalization concept of dependency of attributes, called a *partial dependency* of attributes.

Definition 4. *Let* $S=(U,A,V,f)$ *be an information system and let D and C be any subsets of A. Degree of dependency of attribute D on attributes C, denoted* $C \Rightarrow_k D$*, is defined by*

$$k=\frac{\sum_{X\in U/D}\left|\underline{C}(X)\right|}{|U|}. \qquad (1)$$

Obviously, $0 \le k \le 1$. Attribute D is said to be (totally dependent) depends totally (in a degree of k) on the attribute C if $k = 1$. Otherwise, D is depends partially on C. Thus, attribute D depends totally (partially) on attribute C, if all (some) elements of the universe U can be uniquely classified to equivalence classes of the partition U/D, employing C.

3. The Proposed Technique

3.1. The MDA Technique

In the proposed technique, the rough attributes dependencies in categorical-valued information systems is used to select clustering attribute based on the maximum degree. We have succeed in showing that the proposed technique is able to achieve lower computational complexity with higher purity as compared to the baseline method[12-14]. The proposed technique for selecting partitioning attribute is based on the maximum degree of dependency of attributes. The justification that the higher of the degree of dependency of attributes implies the more accuracy for selecting partitioning attribute is stated in the Proposition 1.

Proposition 1. *Let* $S = (U, A, V, f)$ *be an information system and let D and C be any subsets of A. If D depends totally on C, then*

$$\alpha_D(X) \le \alpha_C(X),$$

for every $X \subseteq U$.

Proof. Let D and C be any subsets of A in information system $S = (U, A, V, f)$. From the hypothesis, we have $IND(C) \subseteq IND(D)$. Furthermore, the partitioning U/C is finer than that U/D, thus, it is clear that any equivalence class induced by $IND(D)$ is a union of some equivalence class induced by $IND(C)$. Therefore, for every $x \in X \subseteq U$, we have $[x]_C \subseteq [x]_D$. And hence, for every $X \subseteq U$, we have

$$\underline{D}(X) \subseteq \underline{C}(X) \subset X \subset \overline{C}(X) \subseteq \overline{D}(X).$$

Consequently

$$\alpha_D(X) = \frac{\left|\underline{D}(X)\right|}{\left|\overline{D}(X)\right|} \le \frac{\left|\underline{C}(X)\right|}{\left|\overline{C}(X)\right|} = \alpha_C(X). \ \square$$

The generalization of Proposition 1 is given below.

Proposition 2. *Let* $S=(U,A,V,f)$ *be an information system and let* $C_1,C_2,\cdots,C_n$ *and D be any subsets of A. If* $C_1 \Rightarrow_{k_1} D, C_2 \Rightarrow_{k_2} D, \cdots, C_n \Rightarrow_{k_n} D$, *where* $k_n \le k_{n-1} \le \cdots \le k_2 \le k_1$, *then*

$$\alpha_D(X) \le \alpha_{C_n}(X) \le \alpha_{C_{n-1}}(X) \le \cdots \le \alpha_{C_2}(X) \le \alpha_{C_1}(X),$$

for every $X \subseteq U$.

Proof. Let $C_1,C_2,\cdots,C_n$ and *D* be any subsets of *A* in information system *S*. From the hypothesis and Proposition 4.1, the accuracies of roughness are given as

$$\begin{gathered} \alpha_D(X) \le \alpha_{C_1}(X) \\ \alpha_D(X) \le \alpha_{C_2}(X) \\ \vdots \\ \alpha_D(X) \le \alpha_{C_n}(X) \end{gathered}$$

Since $k_n \le k_{n-1} \le \cdots \le k_2 \le k_1$, then

$$\begin{gathered} [x]_{C_n} \subseteq [x]_{C_{n-1}} \\ [x]_{C_{n-1}} \subseteq [x]_{C_{n-2}} \\ \vdots \\ [x]_{C_2} \subseteq [x]_{C_1}. \end{gathered}$$

Obviously,

$$\alpha_D(X) \le \alpha_{C_n}(X) \le \alpha_{C_{n-1}}(X) \le \cdots \le \alpha_{C_2}(X) \le \alpha_{C_1}(X).$$

Figure 1 shows the pseudo-code of the proposed technique. The technique uses the dependency of attributes in the rough set theory in information systems. It consists of four main steps. The first step deals with the computation of the equivalence classes of each attribute (feature). The equivalence classes of the set of objects *U* can be obtained using the indiscernibility relation of attribute $a_i \in A$ in information system $S=(U,A,V,f)$. The second step deals with the determination of the dependency degree of attributes. The degree of dependency attributes can be determined using formula in equation (1). The third step deals with selecting the maximum dependency degree. Finally, the attribute is ranked with the ascending sequence based on the maximum of dependency degree of each attribute.

Algorithm: MDA
Input: Dataset without clustering attribute
Output: Clustering attribute
Begin

Step 1. Compute the equivalence classes using the indiscernibility relation on each attribute.

Step 2. Determine the dependency degree of attribute a_i with respect to all a_j, where $i \neq j$.

Step 3. Select the maximum of dependency degree of each attribute.

Step 4. Select a clustering attribute based on the maximum degree of dependency of attributes.

End

Fig. 1. The MDA algorithm

In the proposed technique, it is recommended to look at the next lowest dependencies degree inside the attributes that are tied and so on until the tie is broken.

3.2. Example

The dataset is an animal dataset from Hu[19]. In Table 2, there are nine animals with nine categorical-valued attributes; Hair, Teeth, Eye, Feather, Feet, Eat, Milk, Fly and Swim. The attributes Hair, Eye, Feather, Milk, Fly and Swim have two values. Attributes Teeth has three values, and other attributes have four values.

a. To obtain the dependencies degree of all attributes, the first step of the techniques is to obtain the equivalence classes induced by indiscernibility relation of singleton attributes, i.e., disjoint classes of objects which are contain indiscernible objects.
b. By collecting the equivalence classes, a partition of objects can be obtained. The partitions are shown in Figure 2.
c. The dependency degree of attributes can be obtained using formula in (1). For attribute Hair depends on attributes Teeth, Eye, Feather, Feet, Eat, Milk, Fly and Swim, we have the degrees as shown in Figure 3.

Table 2. Animal world dataset from[19]

Animal	Hair	Teeth	Eye	Feather	Feet	Eat	Milk	Fly	Swim
Tiger	Y	Pointed	Forward	N	Claw	Meat	Y	N	Y
Cheetah	Y	Pointed	Forward	N	Claw	Meat	Y	N	Y
Giraffe	Y	Blunt	Side	N	Hoof	Grass	Y	N	N
Zebra	Y	Blunt	Side	N	Hoof	Grass	Y	N	N
Ostrich	N	N	Side	Y	Claw	Grain	N	N	N
Penguin	N	N	Side	Y	Web	Fish	N	N	Y
Albatross	N	N	Side	Y	Claw	Grain	N	Y	Y
Eagle	N	N	Forward	Y	Claw	Meat	N	Y	N
Viper	N	Pointed	Forward	N	N	Meat	N	N	N

a. $X(\text{Hair} = \text{yes}) = \{1,2,3,4\}$, $X(\text{Hair} = \text{no}) = \{5,6,7,8,9\}$,

$$U / \text{Hair} = \{\{1,2,3,4\},\{5,6,7,8,9\}\}.$$

b. $X(\text{Teeth} = \text{pointed}) = \{1,2,9\}$, $X(\text{Teeth} = \text{blunt}) = \{3,4\}$, $X(\text{Teeth} = \text{no}) = \{5,6,7,8\}$,

$$U / \text{Teeth} = \{\{1,2,9\},\{3,4\}\{5,6,7,8\}\}.$$

c. $X(\text{Eye} = \text{Forward}) = \{1,2,8,9\}$, $X(\text{Eye} = \text{Side}) = \{3,4,5,6,7\}$,

$$U / \text{Eye} = \{\{1,2,8,9\},\{3,4,5,6,7\}\}.$$

d. $X(\text{Feather} = \text{no}) = \{1,2,3,4,9\}$, $X(\text{Feather} = \text{yes}) = \{5,6,7,8\}$,

$$U / \text{Feather} = \{\{1,2,3,4,9\},\{5,6,7,8\}\}.$$

e. $X(\text{Feet} = \text{claw}) = \{1,2,5,7,8\}$, $X(\text{Feet} = \text{hoof}) = \{3,4\}$, $X(\text{Feet} = \text{web}) = \{6\}$, $X(\text{Feet} = \text{no}) = \{9\}$.

$$U / \text{Feet} = \{\{1,2,5,7,8,9\},\{3,4\},\{6\},\{9\}\}.$$

f. $X(\text{Eat} = \text{Meat}) = \{1,2,8,9\}$, $X(\text{Eat} = \text{grass}) = \{3,4\}$, $X(\text{Eat} = \text{grain}) = \{5,7\}$, $X(\text{Eat} = \text{fish}) = \{6\}$.

$$U / \text{Eat} = \{\{1,2,8,9\},\{3,4\},\{5,7\},\{6\}\}.$$

g. $X(\text{Milk} = \text{yes}) = \{1,2,3,4\}$, $X(\text{Milk} = \text{no}) = \{5,6,7,8,9\}$,

$$U / \text{Milk} = \{\{1,2,3,4\},\{5,6,7,8,9\}\}.$$

h. $X(\text{Fly} = \text{no}) = \{1,2,3,4,5,6,9\}$, $X(\text{Fly} = \text{yes}) = \{7,8\}$,

$$U / \text{Fly} = \{\{1,2,3,4,5,6\},\{7,8\}\}.$$

i. $X(\text{Swim} = \text{yes}) = \{1,2,6,7\}$, $X(\text{Swim} = \text{no}) = \{3,4,5,8,9\}$,

$$U / \text{Swim} = \{\{1,2,6,7\},\{3,4,5,8,9\}\}.$$

Fig. 2. The partitions using singleton attributes

$$\text{Teeth} \Rightarrow_k \text{Hair}, \text{ where } k = \frac{\sum_{U / \text{Hair}} \left|\underline{\text{Teeth}}(X)\right|}{|U|} = \frac{|\{3,4\}| + |\{5,6,7,8\}|}{9} = \frac{6}{9}.$$

$$\text{Eye} \Rightarrow_k \text{Hair}, \text{ where } k = \frac{\sum_{U / \text{Hair}} \left|\underline{\text{Eye}}(X)\right|}{|U|} = \frac{|\phi|}{9} = 0.$$

$$\text{Feather} \Rightarrow_k \text{Hair}, \text{ where } k = \frac{\sum_{U / \text{Hair}} \left|\underline{\text{Feather}}(X)\right|}{|U|} = \frac{|\{5,6,7,8\}|}{9} = \frac{4}{9}.$$

$$\text{Feet} \Rightarrow_k \text{Hair}, \text{ where } k = \frac{\sum_{U / \text{Hair}} \left|\underline{\text{Feet}}(X)\right|}{|U|} = \frac{|\{3,4\}| + |\{6\}| + |\{9\}|}{9} = \frac{4}{9}.$$

$$\text{Eat} \Rightarrow_k \text{Hair}, \text{ where } k = \frac{\sum_{U / \text{Hair}} \left|\underline{\text{Eat}}(X)\right|}{|U|} = \frac{|\{3,4\}| + |\{5,7\}| + |\{6\}|}{9} = \frac{5}{9}.$$

$$\text{Milk} \Rightarrow_k \text{ Hair}, \text{ where } k = \frac{\sum_{U/\text{Hair}} \left|\underline{\text{Milk}}(X)\right|}{|U|} = \frac{\left|\{1,2,3,4\}\right| + \left|\{5,6,7,8,9\}\right|}{9} = 1.$$

$$\text{Fly} \Rightarrow_k \text{ Hair}, \text{ where } k = \frac{\sum_{U/\text{Hair}} \left|\underline{\text{Fly}}(X)\right|}{|U|} = \frac{\left|\{7,8\}\right|}{9} = \frac{2}{9}.$$

$$\text{Swim} \Rightarrow_k \text{ Hair}, \text{ where } k = \frac{\sum_{U/\text{Hair}} \left|\underline{\text{Swim}}(X)\right|}{|U|} = \frac{|\phi|}{9} = 0.$$

Fig. 3. The attributes dependencies

Similar calculations are performed for all the attributes. These calculations are summarized in Table 3.

Table 3. The dependencies degree of all attributes from Table 2

Attribute	Degree of dependency							
Hair	Teeth	Eye	Feather	Feet	Eat	Milk	Fly	Swim
	0.666	0	0.444	0.444	0.555	1	0.222	0
Teeth	Hair	Eye	Feather	Feet	Eat	Milk	Fly	Swim
	0	0	0.444	0.444	0.555	0	0.222	0
Eye	Hair	Teeth	Feather	Feet	Eat	Milk	Fly	Swim
	0	0.555	0	0.444	1	0	0	0
Feather	Hair	Teeth	Eye	Feet	Eat	Milk	Fly	Swim
	0.444	1	0	0.444	0.555	0.444	0.222	0
Feet	Hair	Teeth	Eye	Feather	Eat	Milk	Fly	Swim
	0	0.222	0	0	0.555	0	0.2222	0
Eat	Hair	Teeth	Eye	Feather	Feet	Milk	Fly	Swim
	0	0.555	0.444	0	0.333	0	0	0
Milk	Hair	Teeth	Eye	Feather	Feet	Eat	Fly	Swim
	1	0.666	0	0.444	0.444	0.555	0.222	0
Fly	Hair	Teeth	Eye	Feather	Feet	Eat	Milk	Swim
	0.444	0.555	0	0.555	0.444	0.333	0.444	0
Swim	Hair	Teeth	Eye	Feather	Feet	Eat	Milk	Fly
	0	0.222	0	0	0.444	0.333	0	0

With the MDA technique, the first maximum degree of dependency of attributes, i.e. 1 occurs in attributes Hair (Milk), Eye and Feather (i.e., 1) as Table 4 shows. The second maximum degree of dependency of attributes, i.e. 0.666 occurs in attributes Hair. Thus, based on Table 4, attribute Hair is selected as clustering attribute.

3.3. Objects splitting

For objects splitting, we use a divide-conquer method. For example, in Table 2 we can cluster (partition) the animals based on the decision attribute selected, i.e., Hair/Milk. Notice that, the partition of the set of animals induced by attribute Hair/Milk is $\{\{1,2,3,4\},\{5,6,7,8,9\}\}$. To this, we can split the animals using the hierarchical tree as follows.

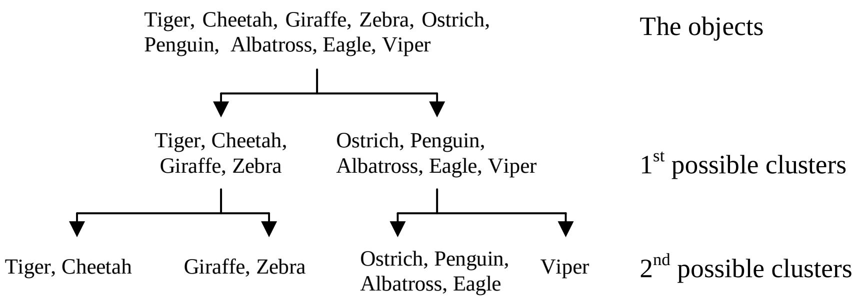

Fig. 4. The objects splitting

The technique is applied recursively to obtain further clusters. At subsequent iterations, the leaf node having more objects is selected for further splitting. The algorithm terminates when it reaches a pre-defined number of clusters. This is subjective and is pre-decided based either on user requirement or domain knowledge.

4. Experiment Test

In order to apply MDA, we use two datasets obtained from the benchmark UCI Machine Learning Repository[18]. We use Lung Cancer and Breast Cancer datasets. The algorithms of MDA for Lung Cancer and Breast Cancer datasets are implemented in MATLAB version 7.6.0.324 (R2008a). They are executed sequentially on a processor Intel Core 2 Duo CPUs. The total main memory is 1 Gigabyte and the operating system is Windows XP Professional SP3.

4.1. Lung Cancer Dataset

The first experiment was conducted on Breast-Cancer-Wisconsin dataset[20]. The data described 3 types of pathological lung cancers. The Authors give no information on the individual variables nor on where the data was originally used. The number of instances is 32, number of attributes is 57 (1 class attribute, 56 predictive), where attribute 1 is the class label. Meanwhile all predictive attributes are nominal, taking on integer values 0-3. Applying MDA on this dataset, we get the values of maximal dependencies among attributes are given in Appendix (refers Table 1 in Appendix A.1). Therefore, we obtain the following clusters.

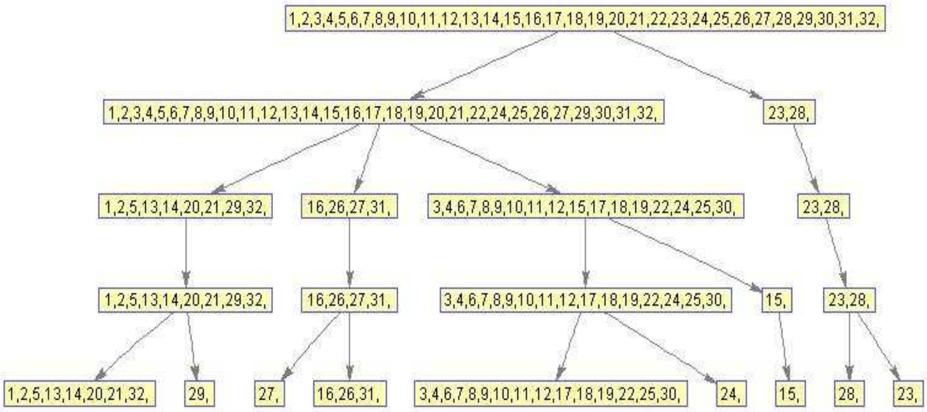

Fig. 5. The lung cancer clusters obtained

4.2. Breast-Cancer-Wisconsin (Original) Dataset

The second experiment was conducted on Breast-Cancer-Wisconsin (original) dataset[21]. The aim of the dataset is to diagnose the breast cancer according to Fine- Needle Aspirates (FNA) test. The dataset was obtained from a repository of a machine-learning database University of California, Irvin. It was compiled by Dr. William H. Wolberg from University of Wisconsin Hospitals, Madison, Madison, WI, United States. It has 10 attributes and 699 records (as of 15 July 1992) with 158 benign and 241 malignant classes, respectively. Applying MDA on this dataset, we get the values of maximal dependencies among attributes are given in Appendix (refers Table 2 in Appendix A.2). Therefore, we obtain the following clusters.

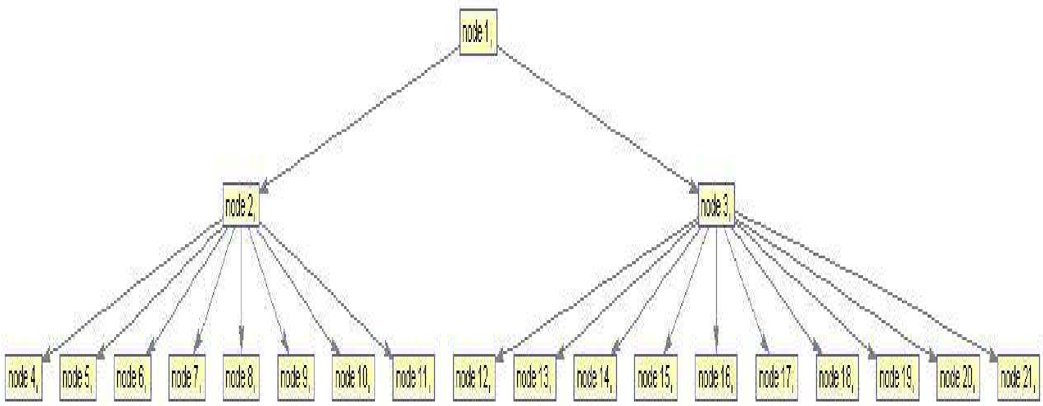

Fig. 6. The breast cancer clusters obtained

The following sets are related to each node in Figure 6.
Node 1 is the set of all objects
Node 2 is the set consist of
1,2,3,4,5,7,8,9,10,11,12,14,17,18,20,23,24,26,27,28,29,30,31,33,34,35,37,44,46,47,60,63,65,68,69,71,74,75,76,77,78,79,80,81,82,87,88,89,90,91,92,93,94,95,96,101,107,109,113,114,115,117,118,119,120,124,126,128,129,130,132,133,134,135,136,137,138,139,141,142,144,145,147,150,151,153,154,157,158,159,160,163,164,165,166,167,171,173,175,176,177,180,184,187,188,189,190,191,192,193,194,197,198,199,202,203,204,207,211,212,214,215,220,223,227,229,234,235,236,237,238,239,242,243,245,249,250,251,252,258,262,264,267,268,269,270,272,273,279,282,283,286,287,290,292,295,296,298,299,300,302,303,306,307,309,311,312,314,318,319,322,324,325,328,329,330,332,333,334,337,338,339,341,342,349,350,351,352,355,356,357,358,359,360,361,362,363,364,365,366,367,369,370,371,372,374,375,376,377,379,380,381,382,383,384,385,386,388,389,390,391,392,393,394,395,396,397,399,401,403,404,405,406,408,409,410,412,414,415,416,417,418,419,420,423,424,425,427,428,429,430,431,432,433,434,436,437,438,440,444,445,446,447,448,449,450,454,455,456,457,458,459,460,461,462,463,464,466,467,470,471,472,476,478,480,481,482,483,484,485,486,487,488,489,490,491,493,494,495,496,497,498,499,502,503,504,506,507,510,511,512,513,514,515,517,518,519,520,521,522,523,524,525,526,527,528,529,530,531,533,534,536,537,538,539,540,541,542,543,544,545,546,547,548,549,550,552,553,558,559,561,562,563,564,565,566,569,570,571,573,575,579,581,582,583,584,585,586,587,588,592,593,595,599,600,601,602,603,604,605,606,607,608,609,610,612,613,614,615,616,617,619,620,622,623,624,625,626,627,628,629,630,631,632,634,635,636,637,638,639,640,641,642,644,645,646,647,648,649,650,651,652,656,657,658,659,660,661,662,663,664,667,668,669,670,671,672,673,674,675,677,678,679,680.
Node 3 is the set consist of
6,13,15,16,19,21,22,25,32,36,38,39,40,41,42,43,45,48,49,50,51,52,53,54,55,56,57,58,59,61,62,64,66,67,70,72,73,83,84,85,86,97,98,99,100,102,103,104,105,106,108,110.
Node 4 is the set consist of
1,5,8,9,10,11,12,17,18,20,23,24,26,27,28,29,30,31,33,34,35,37,44,46,47,63,65,68,69,71,75,76,78,79,80,81,87,88,89,90,91,92,93,94,95,96,101,107,119,124,126,128,129,130,132,133,135,136,137,138,139,142,145,147,151,153,154,157,158,163,164,165,166,167,171,173,175,176,177,180,184,187,188,189,190,192,193,194,197,198,199,202,203,204,207,211,212,214,215,220,223,227,229,235,236,238,242,243,249,250,251,258,262,264,267,268,269,270,272,273,279,282,283,287,290,292,295,296,299,300,302,303,306,309,311,312,314,318,319,322,324,325,328,329,330,332,333,334,337,338,341,342,351,352,355,356,357,358,359,360,361,362,363,364,365,366,367,369,370,371,372,374,375,377,379,380,381,382,383,384,385,386,388,389,391,392,393,394,396,397,399,403,405,408,409,410,412,414,415,416,417,418,419,423,424,425,430,431,432,433,434,436,437,438,440,444,445,446,448,449,450,454,455,456,457,458,459,460,461,462,463,464,466,467,470,472,476,478,481,482,483,484,485,486,487,488,489,490,491,494,495,496,497,498,499,502,503,504,506,507,510,511,512,513,514,515,517,518,519,520,521,522,523,524,525,527,528,529,530,531,533

,534,536,537,538,540,541,542,543,544,545,546,547,548,549,550,552,558,559,561,562,563,564,565,566,569,570,571,573,575,579,581,582,583,584,585,586,587,588,592,593,595,599,600,601,602,603,604,605,607,608,609,613,614,615,616,617,619,620,623,624,625,626,627,628,629,630,631,632,634,636,637,638,639,640,641,642,644,645,646,647,648,649,650,651,652,656,657,658,659,660,661,662,663,664,667,668,669,670,671,672,673,674,675,677,678,680.

Node 5 is the set consist of 3,60,74,82,109,115,118,120,134,144,160,234,239,376,395,404,429,526,606,622,679.

Node 6 is the set consist of 14,77,113,117,150,159,339,349,350,401,406,428,471,553.

Node 7 is the set consist of 4,390,427,493,610,635.

Node 8 is the set consist of 114,141,237,286,298,307,447,480,539,612.

Node 9 is the set consist of 191.

Node 10 is the set consist of 252,420.

Node 11 is the set consist of 2,7,245.

Node 12 is the set consist of 16,36,42,43,56,103,162,172,301,335,422,576,589,621,653.

Node 13 is the set consist of 58,62,64,70,231,413,560,597,633.

Node 14 is the set consist of 13,40,49,59,102,104,257,297,343,443,474,574,618,681.

Node 15 is the set consist of 50,73,84,213,261,266,289,315,345,398,451,475,682.

Node 16 is the set consist of 32,51,54,99,100,140,149,201,210,217,222,280,468,492,508,535,555,654,676,683.

Node 17 is the set consist of 55,97,441,611.

Node 18 is the set consist of 22,25,39,123,225,271,346.

Node 19 is the set consist of 48,52,53,61,85,112,143,181,183,185,206,218,226,247,305,317,336,591,655.

Node 20 is the set consist of 15,45,67,83,108,110,232,241,387.

Node 21 is the set consist of 6,19,21,38,41,57,66,72,86,98,105,106,111,116,121,122,125,127,131,146,148,152,155,156,161,168,169,170,174,178,179,182,186,195,196,200,205,208,209,216,219,221,223.

5. Conclusion

Expert systems and machine learning techniques are gaining popularity in the field of computational biology because of the effective classification and high diagnostic capability. In this paper, we explore MDA technique-an alternative technique for categorical data clustering using rough set theory based on attributes dependencies- for clustering two cancer datasets. These datasets are taken from UCI ML repository. To select a clustering attribute, the maximal degree of the rough attributes dependencies in categorical-valued information systems is used. Further, we use a divide-and-conquer method to partition/cluster the objects. The results show that MDA technique can be used to cluster to the data. Further, we present clusters visualization using two dimensional plot. The plot results provide user friendly navigation to understand the cluster obtained.

Acknowledgement

This work was supported by The Research and Innovations Center, Universiti Malaysia Pahang. The author would like to thank Dr. Iwan Tri Riyadi Yanto for providing Matlab code to cluster the cancer datasets.

References

1. M.C.V. Nascimento, F.M.B. Toledo and A.C.P.L.F. de Carvalho, Investigation of a new GRASP-based clustering algorithm applied to biological data, *Computers & Operations Research*, **37**, 1381–1388 (2010).
2. Z. Huang, Extensions to the k-means algorithm for clustering large data sets with categorical values, *Data Mining and Knowledge Discovery*, **2 (3)**, 283–304 (1998).
3. S. Guha, R. Rastogi and K. Shim, ROCK: a robust clustering algorithm for categorical attributes, *Information Systems*, **25** (5), 345–366 (2000).
4. D. Kim, K. Lee and D. Lee, Fuzzy clustering of categorical data using fuzzy centroids, *Pattern Recognition Letters*, **25 (11)**, 1263–1271(2004).
5. K. Chen and L. Liu, Best K: critical clustering structures in categorical datasets, *Knowledge and Information System*, **20**, 1–33 (2009).
6. H. Zengyou, X. Xiaofei and D. Shengchun, k-ANMI: A mutual information based clustering algorithm for categorical data, *Information Fusion*, **9 (2)**, 223-233 (2008).
7. T. Herawan, I.T.R. Yanto and M.M. Deris, Rough set approach for categorical data clustering. In D. Ślęzak et al. (Eds.): DTA 2009, *Communication of Computer and Information Sciences* Springer-Verlag, **64**, 188–195 (2009).
8. T. Herawan, R. Ghazali, I.T.R. Yanto and M.M. Deris, Rough set approach for categorical data clustering. In a special issue of DTA 2009, *International Journal of Database Theory and Application*, **3 (1)**, 33–52 (2010).
9. Z. Pawlak, Rough sets, *International Journal of Computer and Information Science*, **11**, 341–356 (1982).
10. Z. Pawlak, *Rough sets: A theoretical aspect of reasoning about data*, Kluwer Academic Publisher, (1991).
11. Z. Pawlak and A. Skowron, Rudiments of rough sets, *Information Sciences*. **177 (1)**, 3–27, (2007).
12. T. Herawan, M.M. Deris and J.H. Abawajy, Rough set approach for selecting clustering attribute, *Knowledge Based Systems*. **23 (3)**, 220–231 (2010).
13. T. Herawan and M.M. Deris, A framework on rough set-based partitioning attribute selection, In D.S. Huang et al. (Eds.): ICIC 2009, *Lecture Notes in Artificial Intelligence* Springer-Verlag **5755**, 91–100, (2009).
14. T. Herawan and M.M. Deris, Rough set theory for selecting clustering attribute, International Conference of PCO 2009, *American Institute of Physic*, **1159**, 331–338 (2009).
15. T. Herawan and M.M. Deris, A construction of nested rough set approximation in information systems using dependency of attributes, International Conference of PCO 2009, *American Institute of Physic*, **1159**, 324–331, (2009).
16. T. Herawan, I.T.R. Yanto and M.M. Deris, A construction of hierarchical rough set approximations in information systems using dependency of attributes, In N.T. Nguyen et al. (Eds.): Advances in Intelligent Information and Database Systems, *Studies in Computational Intelligence* Springer-Verlag, **283**, 3–15, (2010).
17. H.L. Chen, Y. Bo, L. Jie and D.Y. Liu, A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis, to appear in *Expert Systems with Applications*, doi:10.1016/j.eswa.2011.01.120 (2011).
18. http://archive.ics.uci.edu/ml/

19. X. Hu, *Knowledge discovery in databases: An attribute oriented rough set approach*, PhD thesis, University of Regina, (1995).
20. http://archive.ics.uci.edu/ml/datasets/Lung+Cancer
21. http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Original)

International Journal of Modern Physics: Conference Series
Vol. 1, No. 1 (2010) 1–5

DOI: 10.1142/insert DOI here

Appendix A.

A.1. Dependency degrees values for Lung Cancer Dataset

Table 1. Dependency degrees values for Lung Cancer Dataset

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.59	0.59	0.88	0.84	0.72	0.81	0.56	0.72	0.09	0.38	0.41	0.53	0.63	0.81	0.81	0.28	0.13	0.13	0.25	0.50	0.25	0.19	0.25	0.25	0.41	0.63	0.31	0.31	0.47
0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.03	0.00	0.06	0.00	0.00	0.00	0.03	0.13	0.00	0.03	0.00	0.00	0.00	0.00	0.09	0.06	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.13	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.25	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.13	0.00	0.00	0.22	0.00	0.09	0.03	0.09	0.19	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.19	0.00	0.09	0.06	0.00	0.00	0.13	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.09	0.06	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.06	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.09	0.03	0.09	0.06	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.31	0.03	0.03	0.53	0.38	0.00	0.19	0.44	0.44	0.66	0.00	0.06	0.28	0.19	0.00	0.03	0.00	0.13	0.25	0.59	0.00	0.00	0.25	0.25	0.06	0.00	0.00	0.31	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.06	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.16	0.06	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.25	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.16	0.03	0.03	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.06	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.22	0.00	0.03	0.03	0.00	0.06	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.13	0.00	0.00	0.00	0.22	0.00	0.09	0.03	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.09	0.06	0.34	0.00	0.19	0.31
0.28	0.03	0.44	0.34	0.00	0.00	0.38	0.22	0.00	0.03	0.03	0.09	0.06	0.28	0.38	0.41	0.03	0.00	0.03	0.00	0.00	0.00	0.00	0.09	0.06	0.34	0.31	0.19	0.00
0.00	0.03	0.03	0.59	0.16	0.00	0.44	0.22	0.28	0.09	0.03	0.09	0.19	0.38	0.00	0.19	0.03	0.00	0.25	0.09	0.00	0.19	0.00	0.09	0.06	0.00	0.00	0.19	0.16
0.28	0.03	0.03	0.25	0.16	0.00	0.00	0.00	0.56	0.09	0.03	0.09	0.06	0.28	0.38	0.00	0.03	0.00	0.00	0.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.16	0.09	0.03	0.00	0.06	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.13	0.00	0.00	0.00	0.00	0.28	0.03	0.03	0.00	0.06	0.38	0.00	0.00	0.03	0.00	0.00	0.03	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.19	0.00
0.00	0.03	0.03	0.00	0.00	0.28	0.00	0.00	0.00	0.09	0.03	0.00	0.06	0.00	0.00	0.00	0.03	0.00	0.13	0.03	0.00	0.00	0.00	0.09	0.06	0.00	0.00	0.19	0.00
0.28	0.03	0.03	0.13	0.16	0.00	0.00	0.00	0.00	0.09	0.03	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.09	0.00	0.00	0.16	0.06	0.00	0.00	0.00	0.16
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.16	0.09	0.03	0.09	0.06	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.06	0.38	0.00	0.00	0.00

0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.03	0.00	0.06	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.16	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.09	0.06	0.00	0.00	0.00
0.00	0.03	0.03	0.13	0.16	0.00	0.00	0.00	0.16	0.03	0.03	0.00	0.00	0.00	0.19	0.19	0.03	0.13	0.00	0.03	0.09	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.03	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.13	0.00	0.00	0.00	0.00	0.00	0.09	0.03	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.34	0.00	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.16	0.09	0.03	0.09	0.06	0.00	0.00	0.00	0.03	0.00	0.13	0.03	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.16	0.09	0.03	0.09	0.06	0.00	0.00	0.00	0.03	0.00	0.13	0.03	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.28	0.03	0.03	0.13	0.16	0.28	0.38	0.00	0.16	0.09	0.03	0.09	0.06	0.28	0.38	0.00	0.03	0.00	0.13	0.03	0.41	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.03	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.09	0.06	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.09	0.06	0.00	0.00	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.03	0.00	0.06	0.00	0.00	0.00	0.03	0.00	0.13	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.03	0.00	0.06	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.03	0.00	0.06	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.25	0.16	0.00	0.00	0.00	0.00	0.09	0.03	0.00	0.00	0.00	0.00	0.19	0.03	0.00	0.00	0.03	0.09	0.00	0.19	0.00	0.09	0.06	0.00	0.00	0.13
0.69	0.03	0.03	0.13	0.16	0.00	0.38	0.34	0.56	0.09	0.03	0.09	0.06	0.28	0.38	0.00	0.03	0.00	0.00	0.03	0.41	0.00	0.00	0.00	0.00	0.00	0.00	0.31	0.19
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.09	0.06	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.09	0.06	0.00	0.00	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.09	0.06	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00
0.69	0.03	0.03	0.25	0.16	0.00	0.38	0.34	0.56	0.09	0.38	0.09	0.19	0.28	0.38	0.19	0.03	0.00	0.00	0.78	0.50	0.00	0.19	0.00	0.16	0.00	0.00	0.31	0.31
0.00	0.03	0.03	0.00	0.00	0.00	0.19	0.00	0.00	0.09	0.03	0.09	0.13	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.13	0.53	0.00	0.19	0.00	0.28	0.03	0.38	0.31	0.06	0.28	0.00	0.00	0.03	0.00	0.13	0.03	0.00	0.00	0.00	0.00	0.25	0.06	0.28	0.00	0.19
0.41	0.03	0.03	0.25	0.16	0.00	0.00	0.00	0.00	0.09	0.03	0.09	0.06	0.00	0.00	0.00	0.03	0.00	0.13	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.38	0.00	0.00
0.69	0.03	0.44	0.25	0.53	0.00	0.56	0.56	0.56	0.09	0.38	0.09	0.19	0.28	0.56	0.59	0.28	0.13	0.13	0.78	0.50	0.00	0.19	0.00	0.25	0.06	0.66	0.00	0.13
0.69	0.03	0.44	0.25	0.53	0.00	0.56	0.56	0.56	0.09	0.38	0.09	0.19	0.28	0.56	0.59	0.28	0.13	0.13	0.78	0.50	0.00	0.19	0.00	0.25	0.06	0.66	0.00	0.13
0.00	0.03	0.03	0.13	0.16	0.00	0.19	0.34	0.28	0.03	0.38	0.31	0.00	0.38	0.00	0.00	0.03	0.00	0.13	0.03	0.09	0.00	0.00	0.00	0.25	0.06	0.28	0.00	0.00
0.28	0.03	0.03	0.25	0.16	0.00	0.38	0.56	0.56	0.09	0.03	0.09	0.06	0.28	0.38	0.41	0.03	0.00	0.13	0.78	0.41	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00
0.00	0.03	0.03	0.13	0.00	0.00	0.00	0.22	0.00	0.09	0.03	0.09	0.06	0.00	0.00	0.00	0.03	0.00	0.13	0.03	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.03	0.09	0.06	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.03	0.09	0.06	0.00	0.00	0.19	0.28	0.13	0.00	0.03	0.00	0.00	0.19	0.00	0.16	0.00	0.00	0.00	0.13
0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.16	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.03	0.13	0.13	0.03	0.00	0.00	0.00	0.00	0.09	0.06	0.00	0.00	0.00
0.00	0.03	0.03	0.13	0.00	0.28	0.00	0.00	0.16	0.03	0.03	0.00	0.47	0.00	0.00	0.41	0.03	0.00	0.00	0.03	0.50	0.00	0.00	0.00	0.09	0.06	0.34	0.00	0.00
0.28	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.16	0.09	0.03	0.09	0.19	0.28	0.00	0.00	0.03	0.00	0.00	0.03	0.09	0.00	0.00	0.00	0.16	0.00	0.00	0.00	0.00

Table 1. Dependency degrees values for Lung Cancer Dataset (*Continued*)

30	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**	**MDA**
0.00	0.00	0.00	0.09	0.00	0.00	0.03	0.13	0.09	0.09	0.19	0.06	0.00	0.13	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.00	0.03	0.03	0.00	0.00	0.00	**0.19**
0.88	0.41	0.41	0.19	0.56	0.50	0.38	0.56	0.44	0.22	0.19	0.53	0.66	0.13	0.25	0.16	0.19	0.06	0.06	0.13	0.13	0.25	0.22	0.22	0.56	0.19	0.72	**0.88**
0.00	0.00	0.00	0.09	0.00	0.00	0.03	0.00	0.09	0.00	0.09	0.00	0.09	0.03	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.00	0.03	0.03	0.00	0.00	0.00	**0.13**
0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.03	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.03	0.03	0.00	0.00	0.00	**0.13**
0.00	0.13	0.13	0.00	0.00	0.00	0.03	0.00	0.09	0.00	0.00	0.00	0.00	0.03	0.06	0.16	0.06	0.06	0.06	0.06	0.06	0.00	0.03	0.03	0.00	0.00	0.00	**0.25**
0.00	0.00	0.00	0.19	0.16	0.00	0.03	0.00	0.03	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.06	0.06	0.13	0.03	0.03	0.00	0.19	0.00	**0.22**
0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.03	0.09	0.09	0.06	0.09	0.03	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.00	0.03	0.03	0.00	0.00	0.00	**0.09**
0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.13	0.09	0.09	0.09	0.06	0.00	0.13	0.06	0.06	0.06	0.06	0.06	0.13	0.13	0.00	0.03	0.03	0.00	0.00	0.00	**0.13**
0.00	0.00	0.00	0.09	0.00	0.00	0.03	0.00	0.09	0.00	0.00	0.00	0.00	0.03	0.06	0.16	0.00	0.00	0.00	0.13	0.06	0.00	0.03	0.03	0.00	0.00	0.00	**0.41**
0.31	0.13	0.13	0.19	0.59	0.13	0.03	0.56	0.66	0.22	0.19	0.06	0.00	0.13	0.25	0.06	0.19	0.06	0.06	0.06	0.13	0.25	0.22	0.22	0.00	0.00	0.28	**0.66**
0.00	0.00	0.00	0.00	0.16	0.00	0.03	0.13	0.09	0.09	0.09	0.06	0.00	0.13	0.06	0.06	0.06	0.06	0.06	0.06	0.13	0.00	0.03	0.03	0.00	0.00	0.00	**0.16**
0.00	0.00	0.00	0.00	0.16	0.00	0.03	0.00	0.03	0.00	0.00	0.00	0.00	0.03	0.06	0.06	0.13	0.00	0.00	0.13	0.00	0.00	0.03	0.03	0.00	0.00	0.00	**0.25**
0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.03	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.03	0.03	0.00	0.00	0.00	**0.06**
0.00	0.00	0.00	0.09	0.00	0.00	0.03	0.00	0.09	0.00	0.00	0.00	0.00	0.03	0.06	0.06	0.00	0.00	0.00	0.06	0.00	0.00	0.03	0.03	0.00	0.00	0.00	**0.16**
0.00	0.00	0.00	0.09	0.00	0.00	0.03	0.00	0.09	0.09	0.00	0.06	0.09	0.03	0.00	0.00	0.06	0.06	0.06	0.06	0.06	0.00	0.03	0.03	0.00	0.00	0.00	**0.09**
0.00	0.00	0.00	0.09	0.00	0.00	0.03	0.00	0.03	0.09	0.00	0.06	0.00	0.13	0.00	0.00	0.06	0.06	0.06	0.06	0.06	0.00	0.03	0.03	0.00	0.00	0.00	**0.22**
0.00	0.00	0.00	0.09	0.16	0.00	0.03	0.00	0.09	0.09	0.00	0.06	0.09	0.03	0.00	0.09	0.06	0.06	0.06	0.06	0.06	0.00	0.03	0.22	0.00	0.00	0.00	**0.34**
0.13	0.28	0.28	0.09	0.16	0.13	0.03	0.44	0.66	0.09	0.00	0.06	0.00	0.03	0.19	0.09	0.06	0.06	0.06	0.06	0.06	0.13	0.03	0.22	0.44	0.00	0.00	**0.66**
0.13	0.41	0.41	0.19	0.16	0.13	0.38	0.00	0.03	0.09	0.09	0.06	0.09	0.03	0.06	0.16	0.19	0.06	0.06	0.13	0.13	0.25	0.03	0.03	0.44	0.00	0.00	**0.59**
0.00	0.00	0.00	0.09	0.00	0.00	0.03	0.00	0.03	0.09	0.00	0.06	0.00	0.03	0.19	0.09	0.06	0.06	0.06	0.06	0.13	0.13	0.03	0.03	0.00	0.00	0.00	**0.56**
0.00	0.00	0.00	0.09	0.00	0.00	0.03	0.13	0.09	0.09	0.09	0.06	0.00	0.13	0.06	0.06	0.06	0.06	0.06	0.13	0.06	0.00	0.03	0.03	0.00	0.00	0.00	**0.16**
0.00	0.41	0.41	0.19	0.16	0.13	0.03	0.00	0.09	0.00	0.00	0.00	0.00	0.03	0.06	0.06	0.13	0.00	0.00	0.06	0.06	0.00	0.03	0.03	0.00	0.00	0.00	**0.41**
0.00	0.28	0.28	0.09	0.16	0.13	0.03	0.00	0.03	0.22	0.00	0.06	0.00	0.13	0.19	0.00	0.06	0.06	0.06	0.00	0.06	0.13	0.03	0.22	0.00	0.00	0.00	**0.28**
0.00	0.00	0.00	0.09	0.16	0.13	0.03	0.13	0.09	0.00	0.09	0.06	0.34	0.13	0.25	0.16	0.13	0.00	0.00	0.06	0.06	0.00	0.22	0.03	0.00	0.00	0.00	**0.34**
0.00	0.00	0.00	0.09	0.00	0.38	0.03	0.13	0.09	0.09	0.09	0.06	0.09	0.13	0.06	0.16	0.06	0.06	0.06	0.13	0.06	0.00	0.03	0.03	0.00	0.00	0.00	**0.38**
0.00	0.00	0.00	0.09	0.00	0.00	0.03	0.00	0.03	0.00	0.00	0.00	0.00	0.03	0.06	0.16	0.00	0.00	0.00	0.13	0.00	0.00	0.03	0.03	0.00	0.00	0.00	**0.16**
0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.09	0.00	0.00	0.00	0.00	0.03	0.00	0.09	0.19	0.06	0.06	0.00	0.06	0.00	0.03	0.03	0.00	0.00	0.00	**0.19**
0.00	0.13	0.13	0.00	0.00	0.00	0.03	0.13	0.09	0.00	0.19	0.06	0.00	0.13	0.06	0.06	0.13	0.00	0.00	0.13	0.06	0.13	0.03	0.03	0.00	0.00	0.00	**0.19**

0.00	0.00	0.00	0.09	0.00	0.00	0.03	0.00	0.03	0.00	0.00	0.06	0.00	0.13	0.00	0.09	0.00	0.00	0.00	0.00	0.06	0.00	0.03	0.03	0.00	0.00	0.00	**0.13**
0.00	0.00	0.00	0.09	0.00	0.00	0.03	0.00	0.09	0.00	0.00	0.00	0.09	0.03	0.00	0.00	0.06	0.06	0.06	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.00	**0.09**
0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.09	0.00	0.00	0.00	0.09	0.03	0.06	0.16	0.06	0.06	0.06	0.06	0.00	0.00	0.03	0.03	0.00	0.00	0.00	**0.34**
0.00	0.13	1.00	0.09	0.00	0.00	0.03	0.00	0.09	0.13	0.00	0.00	0.09	0.03	0.06	0.06	0.06	0.06	0.06	0.13	0.00	0.00	0.03	0.03	0.00	0.00	0.00	**1.00**
0.00	0.13	1.00	0.09	0.00	0.00	0.03	0.00	0.09	0.13	0.00	0.00	0.09	0.03	0.06	0.06	0.06	0.06	0.06	0.13	0.00	0.00	0.03	0.03	0.00	0.00	0.00	**1.00**
0.00	0.13	0.00	0.00	0.00	0.00	0.03	0.00	0.09	0.13	0.00	0.00	0.00	0.03	0.06	0.16	0.00	0.00	0.00	0.13	0.00	0.00	0.03	0.22	0.00	0.00	0.00	**0.41**
0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.03	0.00	0.00	0.00	0.00	0.03	0.06	0.06	0.00	0.00	0.00	0.13	0.00	0.00	0.03	0.03	0.00	0.00	0.00	**0.13**
0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.03	0.00	0.09	0.06	0.00	0.03	0.06	0.06	0.13	0.00	0.00	0.13	0.13	0.00	0.03	0.03	0.00	0.00	0.00	**0.13**
0.00	0.13	0.13	0.13	0.00	0.00	0.00	0.00	0.03	0.09	0.09	0.06	0.09	0.03	0.00	0.00	0.06	0.06	0.06	0.00	0.13	0.13	0.03	0.03	0.00	0.00	0.00	**0.13**
0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.03	0.09	0.00	0.09	0.06	0.00	0.13	0.06	0.16	0.00	0.00	0.00	0.06	0.06	0.00	0.03	0.03	0.00	0.00	0.00	**0.16**
0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.03	0.44	0.09	0.00	0.06	0.09	0.03	0.00	0.09	0.06	0.06	0.06	0.00	0.06	0.00	0.03	0.03	0.00	0.00	0.00	**0.44**
0.16	0.00	0.13	0.13	0.00	0.00	0.00	0.03	0.00	0.03	0.00	0.06	0.34	0.03	0.19	0.09	0.06	0.06	0.06	0.00	0.13	0.00	0.22	0.22	0.00	0.00	0.00	**0.34**
0.00	0.13	0.00	0.00	0.00	0.00	0.13	0.03	0.44	0.03	0.00	0.06	0.00	0.03	0.25	0.16	0.00	0.00	0.00	0.13	0.06	0.00	0.03	0.22	0.00	0.00	0.00	**0.69**
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.03	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.06	0.06	0.00	0.03	0.03	0.00	0.00	0.00	**0.09**
0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.03	0.00	0.09	0.13	0.00	0.00	0.03	0.00	0.00	0.06	0.06	0.06	0.06	0.00	0.00	0.03	0.03	0.00	0.00	0.00	**0.13**
0.00	0.13	0.00	0.00	0.00	0.00	0.13	0.03	0.44	0.03	0.00	0.81	0.53	0.00	0.19	0.09	0.00	0.00	0.00	0.06	0.13	0.13	0.22	0.22	0.00	0.19	0.00	**0.81**
0.00	0.13	0.00	0.00	0.09	0.41	0.00	0.03	0.00	0.03	0.13	0.09	0.00	0.00	0.03	0.16	0.13	0.00	0.00	0.13	0.06	0.13	0.03	0.03	0.00	0.00	0.00	**0.41**
0.00	0.31	0.13	0.13	0.19	0.00	0.13	0.03	0.44	0.59	0.13	0.09	0.00	0.00	0.03	0.25	0.13	0.00	0.00	0.06	0.06	0.13	0.03	0.03	0.00	0.19	0.00	**0.59**
0.00	0.13	0.28	0.28	0.09	0.00	0.38	0.03	0.00	0.09	0.00	0.00	0.00	0.00	0.03	0.19	0.09	0.06	0.06	0.06	0.00	0.00	0.03	0.22	0.00	0.00	0.00	**0.41**
0.69	0.31	0.41	0.41	0.09	0.16	0.38	0.66	0.44	0.66	0.91	0.09	0.47	0.91	0.03	0.19	0.09	1.00	1.00	0.06	0.06	0.75	0.22	0.22	0.44	0.19	0.72	**1.00**
0.69	0.31	0.41	0.41	0.09	0.16	0.38	0.66	0.44	0.66	0.91	0.09	0.47	0.91	0.03	0.19	0.09	1.00	1.00	0.06	0.06	0.75	0.22	0.22	0.44	0.19	0.72	**1.00**
0.00	0.19	0.41	0.41	0.19	0.16	0.13	0.03	0.00	0.03	0.00	0.09	0.47	0.34	0.03	0.25	0.06	0.13	0.00	0.00	0.06	0.13	0.22	0.22	0.00	0.00	0.00	**0.47**
0.00	0.00	0.00	0.00	0.00	0.16	0.50	0.03	0.00	0.09	0.13	0.00	0.06	0.00	0.03	0.19	0.09	0.00	0.00	0.00	0.06	0.75	0.03	0.03	0.00	0.00	0.00	**0.78**
0.00	0.00	0.00	0.00	0.00	0.16	0.38	0.03	0.00	0.09	0.00	0.00	0.06	0.00	0.03	0.00	0.09	0.00	0.00	0.00	0.06	0.13	0.03	0.03	0.00	0.00	0.00	**0.38**
0.00	0.13	0.00	0.00	0.09	0.00	0.13	0.03	0.00	0.09	0.13	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.03	0.00	0.00	0.00	**0.13**
0.16	0.19	0.13	0.13	0.19	0.00	0.13	0.03	0.00	0.09	0.13	0.09	0.00	0.00	0.03	0.00	0.00	0.13	0.00	0.00	0.06	0.00	0.00	0.03	0.00	0.19	0.00	**0.28**
0.00	0.00	0.00	0.00	0.00	0.16	0.00	0.03	0.13	0.09	0.09	0.09	0.06	0.09	0.13	0.06	0.16	0.06	0.06	0.06	0.06	0.06	0.00	0.03	0.03	0.19	0.00	**0.19**
0.00	0.00	0.28	0.28	0.09	0.16	0.13	0.03	0.13	0.09	0.09	0.09	0.06	0.00	0.13	0.06	0.16	0.06	0.06	0.06	0.06	0.06	0.00	0.03	0.22	0.44	0.28	**0.50**
0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.03	0.00	0.09	0.00	0.00	0.00	0.00	0.03	0.25	0.16	0.06	0.06	0.06	0.13	0.06	0.00	0.03	0.03	0.00	0.19	**0.28**

A.2. Dependency degrees values for Breast Cancer Dataset

Table 2. Dependency degrees values for Breast Cancer Dataset

At 1	**At2**	**At3**	**At4**	**At5**	**At6**	**At7**	**At8**	**At9**	**MDA**
0	0	0	0	0	0	0	0	0	**0**
0	0	0	0	0	0	0	0	0	**0**
0	0	0	0	0	0	0	0	0	**0**
0	0	0	0.002928	0	0	0	0	0	**0.002928**
0	0	0	0	0	0	0	0	0	**0**
0	0	0	0	0	0	0	0.004392	0	**0.004392**
0	0	0	0	0	0	0	0	0	**0**
0	0	0	0	0.002928	0	0	0	0	**0.002928**
0	0	0	0	0	0	0	0	0	**0**
0.121523	0.178624	0.095168	0.061493	0.002928	0.019034	0.086384	0.10981	0.04246	**0.178624**